

Article

A Machine-Learning-Based Robust Classification Method for PV Panel Faults

Sufyan Ali Memon ¹, Qaiser Javed ^{2,†}, Wan-Gu Kim ¹, Zahid Mahmood ^{2,†}, Uzair Khan ² and Mohsin Shahzad ^{2,*,†}

¹ Department of Defense Systems Engineering, Sejong University, Seoul 05006, Korea

² Department of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan

* Correspondence: mohsinshahzad@cuiatd.edu.pk

† These authors contributed equally to this work.

Abstract: Renewable energy resources have gained considerable attention in recent years due to their efficiency and economic benefits. Their proportion of total energy use continues to grow over time. Photovoltaic (PV) cell and wind energy generation are the least-expensive new energy sources in most countries. Renewable energy technologies significantly contribute to climate mitigation and provide economic benefits. Apart from these advantages, renewable energy sources, particularly solar energy, have drawbacks, for instance restricted energy supply, reliance on weather conditions, and being affected by several kinds of faults, which cause a high power loss. Usually, the local PV plants are small in size, and it is easy to trace any fault and defect; however, there are many PV cells in the grid-connected PV system where it is difficult to find a fault. Keeping in view the aforesaid facts, this paper presents an intelligent model to detect faults in the PV panels. The proposed model utilizes the Convolutional Neural Network (CNN), which is trained on historic data. The dataset was preprocessed before being fed to the CNN. The dataset contained different parameters, such as current, voltage, temperature, and irradiance, for five different classes. The simulation results showed that the proposed CNN model achieved a training accuracy of 97.64% and a testing accuracy of 95.20%, which are much better than the previous research performed on this dataset.

Keywords: convolutional neural networks; fault detection; photovoltaic cell



Citation: Memon, S.A.; Javed, Q.; Kim, W.-G.; Mahmood, Z.; Khan, U.; Shahzad, M. A Machine-Learning-Based Robust Classification Method for PV Panel Faults. *Sensors* **2022**, *22*, 8515. <https://doi.org/10.3390/s22218515>

Academic Editor: Minyi Xu

Received: 23 September 2022

Accepted: 3 November 2022

Published: 4 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Renewable energy, also known as green energy, comes from natural resources or processes that are constantly replaced; hence, it has less environmental impact compared to fossil fuels. Energy sources have shifted from conventional to renewable in the previous several decades, creating a noticeable paradigm change in power systems. The most common way of using solar energy is from solar panels using the photovoltaic (PV) effect to provide useable electricity [1]. According to the International Renewable Energy Agency (IRENA), during the period from 2010 and 2019, the production of renewable energy grew from 1227 GW to 2537 GW, which resulted in an impressive 106.78% rise [2].

Similarly, PV's power generation has expanded from 40 GW to 580 GW, a phenomenal 1350% increase in capacity from 2010 to 2019 [2]. One of the major advantage of solar energy is its easy accessibility. It is renewable and leaves almost no carbon footprint. Despite the free availability and other attractive qualities, PV systems also face challenges, such as reliability, high initial cost, fault sensitivity, and uncertainty [3]. Energy from renewable sources results in overstressed power transmission networks with compromised power quality [4], a poor voltage profile [5], and increased losses [6,7] due to the fact that these are usually connected to medium- or low-power networks. Similarly, physical, environmental, or electrical circumstances can cause faults in a PV system [8,9].

Figure 1 shows that there are three major faults, which are (i) physical faults, (ii) electrical faults, and (iii) environmental faults. Obviously, the harsh outside environment in which the PV systems are installed makes them sensitive to failures and abnormalities, such as faulty wiring, Open-Circuit (OC) fault, Line-to-Line (LL) fault, Ground Fault (GF), hot spots, dust and snow accumulation, and other environmental impacts. Ultimately, PV arrays have suffered immense failures due to the aforescribed faults, which result in decreased efficiency and a shorter lifespan [10,11]. According to the findings reported in [12], about an 18.9% reduction has been noticed in power generation due to such faults.

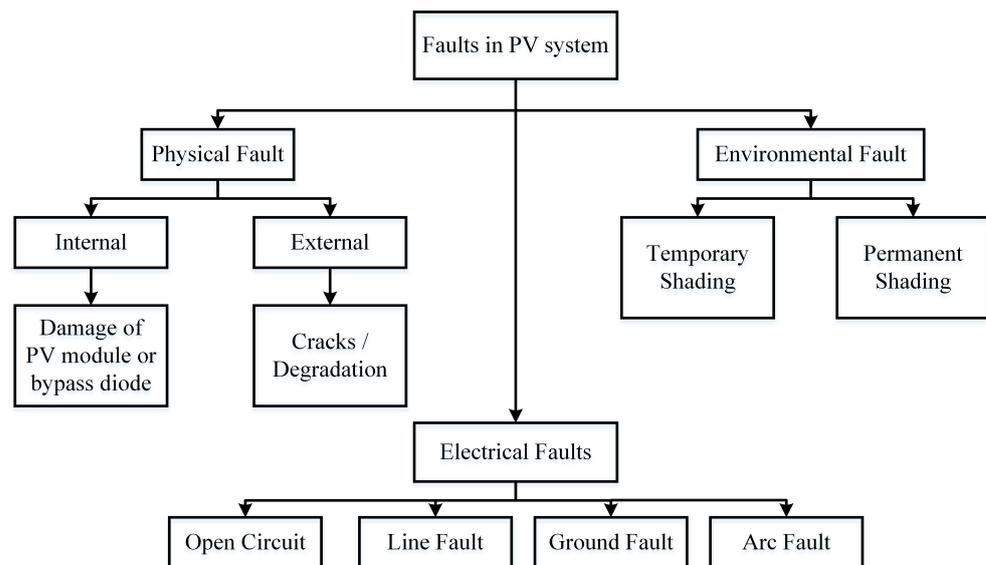


Figure 1. Classification of PV array faults.

To overcome the power loss due to fault occurrence, the National Electric Code (NEC) suggests the usage of Ground Fault Protection Devices (GFPDs), Over-Current Protection Devices (OCPDs), and Arc Fault Circuit Interrupters (AFCIs) to detect LL, LG, and arc faults, respectively. However, the Bakersfield Fire incident in 2009 and that of Mount Holly in 2011 indicate that these devices are unable to identify the error in these specific circumstances [13]. In particular, the nonlinear properties of PV arrays, low irradiance, failure impedance, degradation, and existence of blocking diodes prevent protective devices from tripping under specific situations [13]. Therefore, faults may remain unrecognized for extended periods, demonstrating the limits of standard protection mechanisms in PV arrays. Recent studies, for instance [9,14,15], illustrated the incompatibility of such devices and conventional fault detection techniques. Based on the brief discussion so far, it is evident that fault detection and classification comprise a challenging task with very interesting contributions so far. Therefore, we believe that our work is a novel addition to this domain. The main contributions of this paper are listed below:

- We present a Convolutional-Neural-Network (CNN)-based automatic fault detection and classification method. The proposed machine learning model efficiently reduces power losses in solar PV systems by classifying faults due to its higher accuracy compared to those previously applied.
- We developed an intelligent and robust fault detection and classification technique, which is primarily based on machine learning methods. By using this novel approach, we obtained an increase in fault detection accuracy up to 97.5% from 92.64% compared to the Artificial-Neural-Network (ANN)-based model.
- To the best of our knowledge, the proposed CNN-based model has never been applied to such a big and unbalanced dataset before, for solar system fault detection application. We are optimistic that the proposed method will be a guideline for

beginners and researchers who intend to initiate research in the PV domain using machine-learning-based methods.

The rest of this paper is organized as follows. Section 2 briefly presents recent advancements in fault detection in the PV domain. Section 3 familiarizes the readers with the proposed method. Section 4 presents the simulation results in detail along with the observations and discussions. Finally, Section 5 concludes the paper and provides the possible future work. In each section and to ease the readers' understanding, the abbreviation list shows the common symbols that are used frequently in this paper and their meanings.

2. Related Work

Due to numerous system irregularities, PV systems are frequently sensitive to a range of faults. These irregularities may be temporary or permanent, which ultimately lead to a degradation in system performance. Our study indicates that the Ground Faults (GFs), Short-Circuits (SCs), Open-Circuits (OCs), and shadowing are commonly occurring faults in a PV system. However, this paper considers four major and frequently occurring faults in the PV system, which are the OC, the SC, partial shading, and degradation. The schematic structure of these faults is shown in Figure 2. Below, we briefly discuss these faults:

- **The OC faults:** An OC fault in PV arrays is a disconnection issue inside a string or between two nearby strings [16]. It can occur for a variety of reasons, including a broken cable connecting two strings, an object falling on the panels, or a weak contact between two points [16].
- **The SC faults:** SC faults are caused by an accidental connection between two PV array points with potential variable values. It can occur within the same string or between two adjacent ones [17].
- **Partial shading:** Partial shading is the situation where the PV modules are partially shaded, not receiving proper sunlight to produce energy. It is a temporary situation that causes a decrease in output power [15]. There are two categories of shading [18]. The first is static shading, whereas the second is dynamic shading. Static shading is created by the accumulation of dust, leaves, and bird droppings on the glass, whereas dynamic shading is caused by a momentary shadow cast by surrounding buildings or trees.
- **Degradation:** PV systems are susceptible to degradation owing to front surface soiling, optical degradation due to continuous exposition to UV light, a rise in series resistance or a decrease in shunt resistance, a reduction in the SC current, etc. [19,20]. This error may be module-specific or occur over the entire array over time, resulting in a decline in system performance.

In [13], a fault detection model using the neural network was proposed for PV cells to detect various faults, such as the LL, the OC fault, and partial shading. The authors reported a 73.53% fault detection accuracy. In [21], a novel approach to detect microcracks using short- and long-term deep features and the Deep Convolutional Neural Network (DCNN) was developed. This method is an encouraging solution to analyze faults that appear on solar cell surfaces. In [22], the authors created a fused multi-channel CNN to identify solar cell surface imperfections. In this work, infrared and the neural networks were used to locate and construct an autonomous and robust fault detection system. In [23], faults frequently occurring in the PV domain were investigated using the CNN. To identify and characterize problems in the P modules, such as dust, shadows, and breakage, the CNN and RGB pictures were employed in [24]. This work primarily integrated the RGB images with the CNN-based model to yield encouraging results. Thermal pictures of the PV modules were classified using the SVM-based model into three categories, which were healthy, defective, and hotspot [25]. The works presented in [21–25] are nice efforts to detect and classify various faults in different environments and weather condition. However, few of these works employed image-processing-based methods. The pixelwise operations to detect and classify various faults consume relatively more time than a few of the methods

compared therein. Moreover, image processing methods pose a major threat to developing a real-time fault detection system.

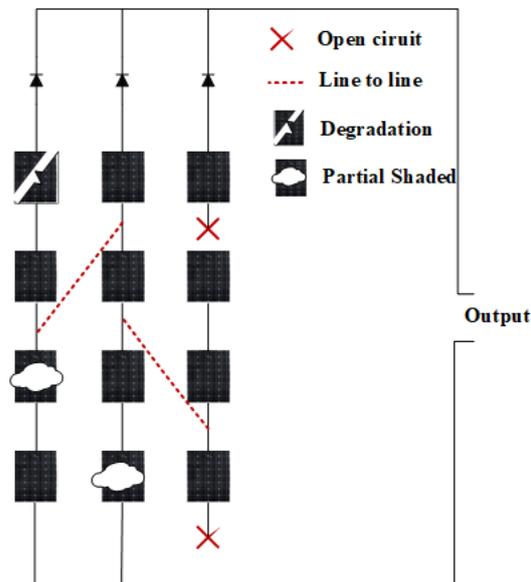


Figure 2. Schematic structure of different faults.

In [26], the researchers developed an SVM-based method to classify the OC and the LL faults, whereas [27,28] investigated the classification of LL faults only. These models reported good accuracy; however, the SVM can be used for binary classification and on simple datasets only. Moreover, the results of the SVM were not as effective on heavy multiclass datasets.

Recently, a few researchers developed the ANN-based fault detection model for various environments [29,30]. In particular, the accuracy of [29] was 92.64%, which is insufficient for a sensitive field. Reference [30] utilized a fairly simple dataset, which can be quickly categorized by simple classification algorithms such as the SVM and decision trees. In [31], a probabilistic-neural-network-based monitoring system was developed to detect the LL and OC faults in a 1.8 kW PV system. In their work, the I-V curve was used as an approach to distinguish among various faults. Their results were encouraging, and the probabilistic based method was one of the best performers for the ANN in the numerical classification technique. In [32], the authors also used the probabilistic method to detect the OC, the SC, the GF, and the hotspots. In [33], the developed model was designed for two different cases of faults, which were (i) single fault occurrence, which included the OC, the LL, and the shading, and (ii) multi-fault occurrence at the same time. The model was tested on a 1.22 kW PV array at the university of NCEPU, China. In [34], twelve different cases of fault were classified through the PNN, on a heavy dataset, which was collected from a 10 kW PV plant and utilized for training and testing. The authors reported a fair accuracy of 92.48% fault detection.

3. Proposed Method

This section describes in detail our proposed fault detection and classification method. Numerous approaches for data normalization, training, validation, and testing of neural networks have been presented by researchers. Figure 3 shows the flowchart of the proposed methodology. While developing our algorithm, we utilized the concepts from the CNN domain.

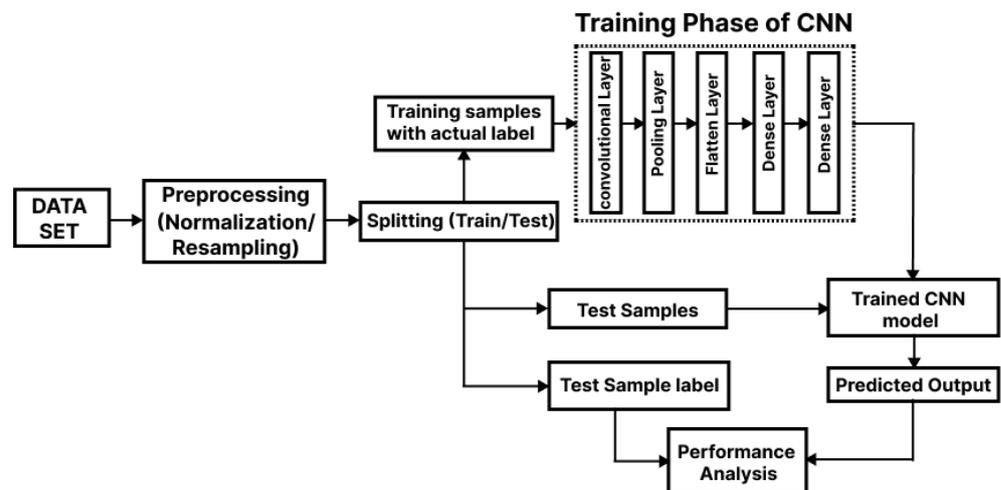


Figure 3. Proposed system architecture for fault diagnosis.

From the literature review, we observed that the CNN is one of the most-popular ANN architectures. The CNN is specifically used in image classification, object detection, and various other computer vision and machine learning tasks. Moreover, the CNN is also extensively used in various other domains, such as Natural Language Processing (NLP) and recommender systems [35,36]. Since we intend to automate the fault detection and classification phenomenon, one of the main reasons for us to use the CNN is that it automatically extracts useful features from the input data without any manual intervention [37].

As shown in Figure 3, the test dataset was initially preprocessed by through normalization and resampling procedures, followed by a splitting phase. In the later stage, both the test samples and test samples of the labeled faults are fed to the trained CNN module, which yields the fault classification prediction. It is worth mentioning here that the CNN architecture shown in Figure 4 used in our fault detection method comprises different layers, which include three major layers, which are the convolutional layer, pooling layer, and fully connected layer.

As shown in Figure 4, the convolutional layer is just the multiplication and accumulation process of our input signal $x = [x_0, x_1, x_2, x_3, \dots, x_n]$ and the kernel $k = [k_{-p}, k_{-p+1}, \dots, k_0, \dots, k_{p-1}, k_p]$. The kernel is slid over the whole input to execute the convolutional operation, which generates a feature map. The convolutional operation used in our can be expressed using Equation (1).

$$f_i = \sum_{j=-p}^p x_{i-j}k_j \quad (1)$$

where f_i is the feature extracted from the input signal x .

To learn some useful nonlinear patterns from the input, a nonlinear activation function was applied to the model. Different activation functions, such as sigmoid and tanh, are also used. However, the most commonly used is the Rectified Linear Unit (ReLU), which is expressed by Equation (2).

$$y = \max(0, x_i) \quad (2)$$

As shown in Figure 4, after convolution, pooling layers are added, which help reduce the spatial dimension of the representation to minimize the computations. This also helps to reduce the number of parameters, which in later stages helps with training time reduction. Normally, two functions are used in the pooling operation, which are average pooling and max pooling. In our model, we used max pooling. The outputs of the pooling layers are fed to the Fully Connected (FC) layer, which is the mandatory layer of the ANN. In the CNN, the convolutional layer and pooling layer extract the features from the data, whereas the FC layer performs the classification.

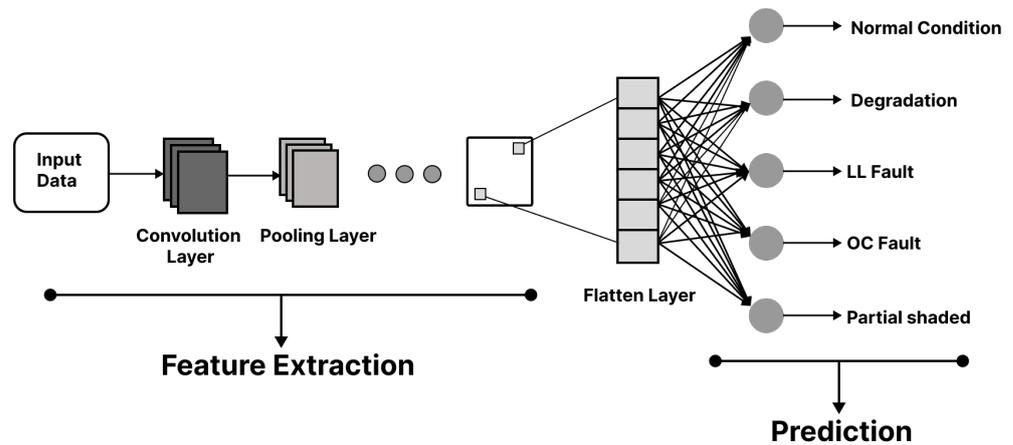


Figure 4. The CNN architecture.

The SoftMax function is used as the activation function for the final classification in the output layer and computes the probability values of all the classes using Equation (3).

$$P(\vec{Z})_i = \frac{e^{Z_i}}{\sum_{j=1}^J e^{Z_j}} \quad (3)$$

where Z_i is the input vector of the SoftMax function. Different range $P(\vec{Z})_i$ values were selected according to the number of classes.

Algorithm 1 illustrates the CNN model designed for the classification of fault occurrence in the PV panel. As shown in lines (2)–(3) of Algorithm 1, firstly, the dataset is preprocessed through normalization and resampling, then it passes through a convolution filter of size 6×1 , which extracts the useful features, and that feature extraction results in the feature map. The extracted feature F is then passed through the MaxPool filter of size 2×1 shown in line (4) of Algorithm 1. Further, as shown in line (5) of Algorithm 1, the output of the pooling layer is converted into a vertical vector by the flatten layer. The aforementioned data are now processed by the fully connected layer as shown in line (6), where advanced features and the probabilities of all classes are computed through the dense layers using 16 nodes. Finally, as shown in line (7), the output layer with 5 nodes is applied. The 5 nodes, which yield the classification, are normal condition, degradation, the LL faults, the OC faults, and partial shading.

Algorithm 1 CNN algorithm for PV fault detection.

1. **Input:** $[X,y]$; label raw data.
 2. **Preprocessing:** normalization and resampling
 3. **Convolution layer:** $CL \leftarrow [X,y]$ raw data sent to the CNN for feature extraction
 F : feature vector (feature map) extracted from convolution layer
 4. **Pooling layer:** $MaxPool \leftarrow F; [X,y]$ downsampling feature map with MaxPool F_D ;
downsampled feature vector or feature map
 5. **Flatten layer:** $F_L \leftarrow F_D$; change dimension of F_D
 F_L : flatten feature map
 6. **Fully connected layer:** $F_C \leftarrow F_L$; compute more advanced features and give
probability values of each class
 P : probabilities of all the classes
 7. **Output:** classifies the test data
-

Now that the algorithm has been developed, in the next section, we discuss in detail the simulation results along with our findings.

4. Simulation Results

We performed detailed simulations using the Google Colab platform, which provides significant and substantial resources in the form of the Google Compute Engine with 12 GB of RAM and 107 GB of disk storage. Before we proceed further, below, we briefly describe the datasets that were used in our work.

4.1. Dataset Description

Dataset-1: This dataset was developed by Lazzaretti et al. [29] in 2020. This dataset contains different features, such as current, voltage, temperature, and irradiation level. The dataset contains these features for five distinct conditions, which are the OC, the LL, partial shading, degradation, and normal condition. In addition, in our work, we selected the following six PV features as the input to train the model:

- x_1 : voltage of string 1;
- x_2 : voltage of string 2;
- x_3 : current of string 1;
- x_4 : current of string 2;
- x_5 : irradiation level;
- x_6 : temperature of PV panel;
- x_7 : fault label.

Figure 5 shows the comparisons of the power produced during a normal condition and four other faulty conditions. It is obvious from the green line in Figure 5 that the power produced under normal circumstances is much greater than that under faulty conditions. In the case of shadowing, the PV panel produces the least current. As a result, the power produced is less than 10 Watts. For other faults, such as short-circuit, degradation, and open-circuit faults, as shown in Figure 5, the amount of produced power is well below the desired level.

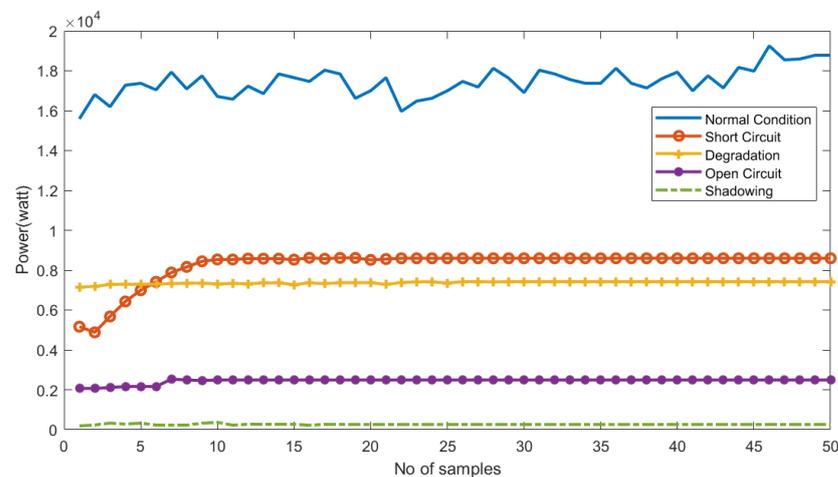


Figure 5. The PV power production under normal and faulty conditions.

4.2. Dataset Preprocessing

In our work, the dataset was preprocessed before being fed to the CNN module. To accelerate the learning procedure and ensure fast convergence, the dataset was initially normalized using standard normalization techniques, as shown by Equation (4).

$$z = \frac{x - u}{s} \quad (4)$$

where z is the normalized value replacement for x , u is the sample's mean, and s is the standard deviation of each feature. After normalization, oversampling was performed as the dataset was imbalanced, because the normal condition contained 886,884 samples, whereas

the SC class contained only 1228 samples [38]. Later, the dataset was randomly partitioned into two subsets, which were the training dataset (80%) and testing dataset (20%).

4.3. The CNN Performance Analysis

To analyze the performance of a neural network or any other machine learning model, learning curves are the most widely used tools. A learning curve is the plot of the model performance based on experience. Looking at the learning curves, one can easily say how well the model finds a relation between the input and output, as well as whether the model is overfit, underfit, or the best-fit model. Usually, we can find two different learning curves.

Training curves: These curves indicate an estimation of how well the model is learning over time from the training dataset.

Validation curves: These curves indicate an estimation of how well the model is generalizing based on how it behaves for unseen data.

The purpose of training a machine learning model is to find the best-fit model.

The best-fit model is identified by the validation and training losses, which decrease the stability point along with a minimal space between the two last loss values. A typical loss function is cross-entropy. The loss function in our model is defined by Equation (5).

$$loss = -\frac{1}{N} \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (5)$$

where N is the total number of examples, y_i is the actual value, and \hat{y}_i is the predicted value. The losses' learning curves for the dataset are shown in Figure 6. Clearly, it can be seen in Figure 6 that, for the epochs set to 50, both the training and validation curves fit best in the given range.

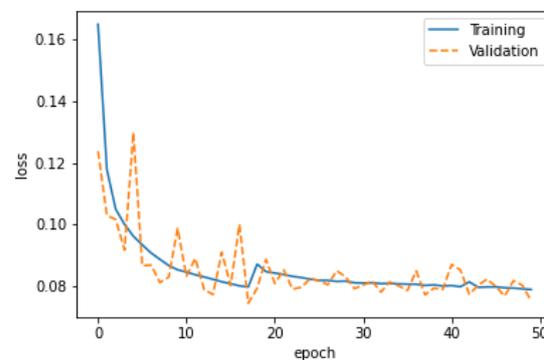


Figure 6. Best-fit model.

4.4. Accuracy Curves

The accuracy parameter is frequently used to evaluate a model's performance. It is defined as the ratio of correct predictions to all predictions. In our work, the accuracy of a model is calculated by using Equation (6).

$$Accuracy = \frac{TN + TP}{TN + FN + FP + TN} \quad (6)$$

where the term TN stands for True Negatives and indicates the number of negatively classified cases that were correctly identified. Similarly, TP denotes True Positives and depicts the quantity of correctly identified positive cases. The term FP denotes the number of False Positives that were mistakenly categorized as positive, while FN denotes the number of False Negatives that were mistakenly classified as negative. Figure 7 shows the training and validation accuracy of our proposed CNN-based model. It can be observed from Figure 7 that, up to 20 epochs, the validation curve tried to match the training curve. After 40 epochs, the difference between the curves was minimized, which further indicates

that the model was well trained at 50 epochs. Observe that the difference between the training curves and validation curves is the basic way to find a model's behavior. In our case, the difference was much less, which shows that the model will predict well for unseen data.

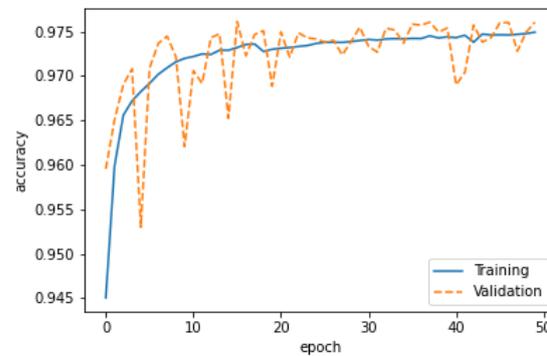


Figure 7. Training and validation accuracy.

4.5. Confusion Matrix

The confusion matrix shows the summary of how a model performs on new data, which is usually the test data. The confusion matrix of our model is shown in Figure 8. In Figure 8, the x-axis shows the predictions and the y-axis shows the actual labels of the classes.

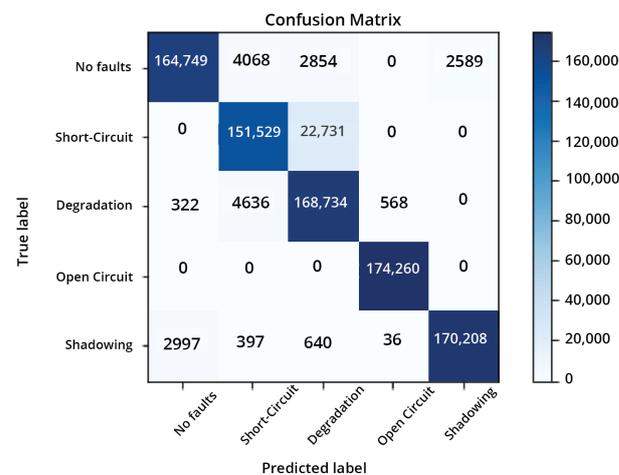


Figure 8. Confusion matrix.

The accuracy achieved for our model for the normal condition, the SC, degradation, the OC, and shadowing was 94.54%, 86.95%, 96.82%, 100%, and 97.67%, respectively. The total testing accuracy of the model was 95.20%.

4.6. Comparison

From the aforementioned analysis, we observed that our model was well trained and yielded encouraging training and testing accuracy. In this section, we compare our work with four recent works. Table 1 shows the comparison of our work with other models that were recently published on fault detection in the PV domain. For a fair comparison, we also performed detailed experiments on a multi-class fault dataset, which we briefly describe in Section 4.1. A few of the important observations from Table 1 are listed below:

- It can be seen in Table 1 that, among all the compared methods, the work reported in [13] yielded the least accuracy of 73.53% to classify the multi-class faults, such as the LL, the OC, and partial shading faults. This was due to the fact that, in their method, the last few layers of a pre-trained AlexNet are fine-tuned to produce several types of

outputs. Similarly, the authors also obtained features from the fully connected (fc7) layer of a pre-trained AlexNet and then used this in conjunction with classical ML methods for classification.

- The works reported in [26,32] yielded the same 97% fault detection accuracy. In particular, the developed method in [26] uses the SVM framework to classify only the LL and the OC faults, whereas Reference [32] successfully classified the GF, the OC, the SC, and hotspot faults by utilizing the PNN framework to yield a similar accuracy.
- Moreover, the authors of [29] also achieved a 92.64% fault detection accuracy for the multi-class faults, such as the OC, the SC, degradation, and shadowing. This is an interesting solution in the fault detection domain, where the authors introduced a recursive linear model to detect faults in the system, primarily through the use of irradiance on the PV panel as the input signals and power as the output.
- Similarly, the work published in [33], utilized a probabilistic framework to classify various faults and, thereby, yielded a good accuracy of 94.69%. This model is also a novel contribution to the fault detection domain and achieved encouraging results by employing several sequential steps. Initially, the authors analyzed the performance of seven indicators to accurately predict the nonlinear output behavior of the PV system under changing environmental conditions. Moreover, several fault cases, considering single-fault types and another three faults considering multiple fault types, were also investigated. In the final stages of this sequential algorithm, the typical fault types were classified and detected using sequential probabilistic neural network models, which gave an effective classification of the data inputs.
- The aforesaid works are sound solutions in the PV domain to reliably detect and classify various faults. However, as seen in Table 1, the proposed fault detection model yielded the highest accuracy among all the compared works. Moreover, our developed CNN-based fault detection model yielded much better accuracy than [13], who also used the CNN framework. On the multi-class dataset developed by [29], we were able to improve the multi-class fault detection accuracy by 5% and 24.11% compared to [13]. Our study indicates that the higher fault detection accuracy in our work was made possible due to the intelligent utilization of the Application Program Interfaces (APIs), which encouraged us to design and operate different layers of the CNN model. Moreover, Google Tensorflow perfectly created and managed these APIs. Furthermore, our proposed method allows real-time fault detection and classification. The training samples along with the actual labels were fed to the training phase of the CNN, which involved various layers, such as convolutional, pooling, flatten, and dense layers, resulting in the intelligently trained CNN model. Later, this model handled the test samples effectively.

Table 1. Accuracy comparison of our model with previous work.

Ref.	Model	Faults under Consideration	Accuracy
[13]	CNN	LL, OC, partial shading	73.53%
[26]	SVM	LL, OC	97%
[29]	ANN	OC, SC, degradation, shadowing	92.64%
[33]	PNN	GF, OC, SC, hotspot	97%
Proposed Model	CNN	OC, SC, degradation, shadowing	97.64%

4.7. Computational Complexity

Figure 9 shows the detailed computational complexity in terms of the time consumed to detect and classify various faults. As can be seen in Figure 9, our model took a one-time training time of 58 min and 18 s for 50 epochs, and this can vary according to the number of epochs set for the simulation; it can also be affected by the RAM of the PC on which the model is trained. For fault detection, such as for the Normal Condition (NC), it consumed 0.16 s. For the SC faults, it took 0.09 s. For degradation and Open-Circuit (OC), only 0.08 s

was used to detect and classify these faults. Finally, for shadowing, our proposed method took 0.07 s to detect and classify this fault.

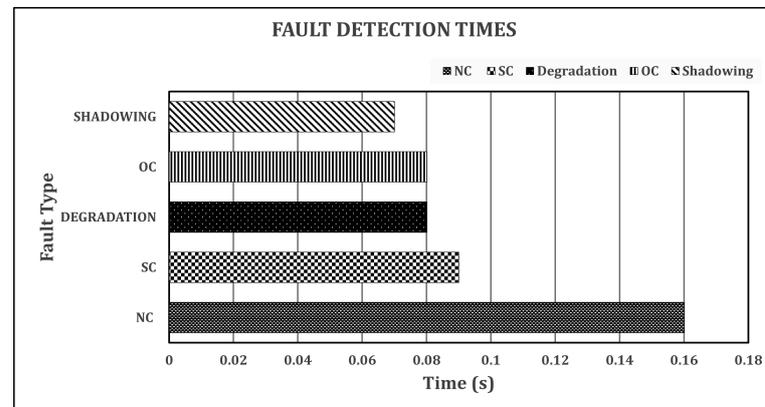


Figure 9. Fault detection times.

5. Discussion

Although the aforescribed analysis shed detailed light on the performance of the developed machine-learning-based fault detection method, for the readers' more in-depth information, below, we briefly discuss the following important points:

- Our study indicated that there is a severe lack of standard protocols to generate and analyze various faults. Similarly, there is a scarcity of available diverse public datasets. We are optimistic that future studies will present more diverse datasets to detect, identify, and classify various faults.
- Our proposed fault detection model classified five different types of fault cases efficiently. However, we observed that the proposed CNN-based fault detection model is heavy compared to other machine algorithms, such as the Discrete Fourier Transform (DFT) and the SVM. Moreover, our proposed model presented a comparable performance to the other works listed in Table 1.
- Our proposed model works in a near real-time fault detection and classification manner. Similarly, our findings suggest that shadowing events are caused by real shadowing, which makes them a challenging task to be accurately characterized. Similarly, controlled shadowing normally increases the performance for a class.

6. Conclusions

Maintaining continuous energy production in PV systems is a critical issue for power utilities, which has attracted much attention from the academic community. Various methods are appearing every day to automate and mitigate the production deviations in PV plants. PV systems are susceptible to various faults and breakdowns. An early fault diagnosis is critical for the PV system's effectiveness and reliability. We presented a CNN-based fault detection algorithm for PV arrays to properly distinguish failure categories. This model was trained on a heavy dataset. The dataset was preprocessed before being fed to the CNN module. The proposed model was well trained and yielded a training accuracy of 97.64%, as well as a validation accuracy of 97.67%.

In the future, we intend to modify the CNN architecture of the model to further improve the detection and classification accuracy for multiple faults' detection. Moreover, we also aim to optimize the proposed model to be reliably used in other fields, such as fault detection in wind turbines.

Author Contributions: Conceptualization, Q.J., M.S. and S.A.M.; methodology, Q.J., M.S. and Z.M.; software, Q.J., Z.M. and M.S.; validation, U.K. and M.S.; formal analysis, Z.M., U.K. and M.S.; investigation, W.-G.K., S.A.M. and U.K.; resources, W.-G.K., Z.M., U.K. and M.S.; data curation, W.-G.K. and U.K.; writing—original draft preparation, M.S., Q.J. and Z.M.; writing—review and editing,

W.-G.K., Z.M. and M.S.; supervision, M.S.; project administration, M.S. and Z.M.; funding acquisition, S.A.M. and W.-G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00166977).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be provided upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFCI	Arc Fault Circuit Interrupter
ANN	Artificial Neural Network
API	Application Programming Interface
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DT	Decision Tree
FC Layer	Fully Connected Layer
GF	Ground Fault
GFPD	Ground Fault Protection Device
GW	Gigawatt
IREA	International Renewable Energy Agency
KW	Kilowatt
LL	Line-to-Line
NEC	National Electric Code
NLP	Natural Language Processing
OC	Open-Circuit
OCPD	Over-Current Protection Device
PV	Photovoltaic
PNN	Probabilistic Neural Network
RGB	Red Green Blue
SC	Short-Circuit
SVM	Support Vector Machine
UV	Ultraviolet

References

1. Antonanzas, J.; Quinn, J. Net environmental impact of the PV industry from 2000–2025. *J. Clean. Prod.* **2021**, *311*, 127791. [[CrossRef](#)]
2. Adhya, D.; Chatterjee, S.; Chakraborty, A.K. Performance assessment of selective machine learning techniques for improved PV array fault diagnosis. *Sustain. Energy Grids Netw.* **2022**, *29*, 100582. [[CrossRef](#)]
3. Bosman, L.B.; Leon-Salas, W.D.; Hutzler, W.; Soto, E.A. PV system predictive maintenance: Challenges, current approaches, and opportunities. *Energies* **2020**, *13*, 1398. [[CrossRef](#)]
4. Shahzad, M.; Shafiullah, Q.; Akram, W.; Arif, M.; Ullah, B. Reactive power support in radial distribution network using mine blast algorithm. *Elektron. Elektrotech.* **2021**, *27*, 33–40. [[CrossRef](#)]
5. Shahzad, M.; Gawlik, W.; Palensky, P. Voltage Quality Index based method to quantify the advantages of optimal DG placement. In Proceedings of the 2016 Eighteenth International Middle East Power Systems Conference (MEPCON), Cairo, Egypt, 27–29 December 2016; pp. 759–764.
6. Shahzad, M.; Akram, W.; Arif, M.; Khan, U.; Ullah, B. Optimal siting and sizing of distributed generators by strawberry plant propagation algorithm. *Energies* **2021**, *14*, 1744. [[CrossRef](#)]
7. Shahzad, M.; Ahmad, I.; Gawlik, W.; Palensky, P. Active power loss minimization in radial distribution networks with analytical method of simultaneous optimal DG sizing. In Proceedings of the 2016 IEEE International Conference on Industrial Technology (ICIT), Taipei, Taiwan, 14–17 March 2016; pp. 470–475.
8. Ji, D.; Zhang, C.; Lv, M.; Ma, Y.; Guan, N. Photovoltaic array fault detection by automatic reconfiguration. *Energies* **2017**, *10*, 699. [[CrossRef](#)]

9. Pillai, D.S.; Rajasekar, N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew. Sustain. Energy Rev.* **2018**, *91*, 18–40. [[CrossRef](#)]
10. Chen, Z.; Wu, L.; Cheng, S.; Lin, P.; Wu, Y.; Lin, W. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics. *Appl. Energy* **2017**, *204*, 912–931. [[CrossRef](#)]
11. Kim, J.; Rabelo, M.; Padi, S.P.; Yousuf, H.; Cho, E.C.; Yi, J. A review of the degradation of photovoltaic modules for life expectancy. *Energies* **2021**, *14*, 4278. [[CrossRef](#)]
12. Appiah, A.Y.; Zhang, X.; Ayawli, B.B.K.; Kyeremeh, F. Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques. *Int. J. Photoenergy* **2019**, *2019*. [[CrossRef](#)]
13. Aziz, F.; Haq, A.U.; Ahmad, S.; Mahmoud, Y.; Jalal, M.; Ali, U. A novel convolutional neural network-based approach for fault classification in photovoltaic arrays. *IEEE Access* **2020**, *8*, 41889–41904. [[CrossRef](#)]
14. Pillai, D.S.; Rajasekar, N. An MPPT-based sensorless line–line and line–ground fault detection technique for PV systems. *IEEE Trans. Power Electron.* **2018**, *34*, 8646–8659. [[CrossRef](#)]
15. Kumar, B.P.; Ilango, G.S.; Reddy, M.J.B.; Chilakapati, N. Online fault detection and diagnosis in photovoltaic systems using wavelet packets. *IEEE J. Photovoltaics* **2017**, *8*, 257–265. [[CrossRef](#)]
16. Kase, R.; Nishikawa, S. Fault detection of bypass circuit of PV module—Detection technology of open circuit fault location. In Proceedings of the 2016 19th International Conference on Electrical Machines and Systems (ICEMS), Chiba, Japan, 13–16 November 2016; pp. 1–4.
17. Pillai, D.S.; Blaabjerg, F.; Rajasekar, N. A comparative evaluation of advanced fault detection approaches for PV systems. *IEEE J. Photovoltaics* **2019**, *9*, 513–527. [[CrossRef](#)]
18. Shenoy, K.L.; Nayak, C.G.; Mandi, R.P. Effect of partial shading in grid connected solar pv system with fl controller. *Int. J. Power Electron. Drive Syst.* **2021**, *12*, 431. [[CrossRef](#)]
19. Piliougine, M.; Guejia-Burbano, R.; Petrone, G.; Sánchez-Pacheco, F.; Mora-López, L.; Sidrach-de Cardona, M. Parameters extraction of single diode model for degraded photovoltaic modules. *Renew. Energy* **2021**, *164*, 674–686. [[CrossRef](#)]
20. Noman, M.; Tu, S.; Ahmad, S.; Zafar, F.U.; Khan, H.A.; Rehman, S.U.; Waqas, M.; Khan, A.D.; Rehman, O.U. Assessing the reliability and degradation of 10–35 years field-aged PV modules. *PLoS ONE* **2022**, *17*, e0261066. [[CrossRef](#)]
21. Qian, X.; Li, J.; Cao, J.; Wu, Y.; Wang, W. Micro-cracks detection of solar cells surface via combining short-term and long-term deep features. *Neural Netw.* **2020**, *127*, 132–140. [[CrossRef](#)]
22. Zhang, X.; Hao, Y.; Shangguan, H.; Zhang, P.; Wang, A. Detection of surface defects on solar cells by fusing Multi-channel convolution neural networks. *Infrared Phys. Technol.* **2020**, *108*, 103334. [[CrossRef](#)]
23. Alves, R.H.F.; de Deus Júnior, G.A.; Marra, E.G.; Lemos, R.P. Automatic fault classification in photovoltaic modules using Convolutional Neural Networks. *Renew. Energy* **2021**, *179*, 502–516. [[CrossRef](#)]
24. Espinosa, A.R.; Bressan, M.; Giraldo, L.F. Failure signature classification in solar photovoltaic plants using RGB images and convolutional neural networks. *Renew. Energy* **2020**, *162*, 249–256. [[CrossRef](#)]
25. Ali, M.U.; Khan, H.F.; Masud, M.; Kallu, K.D.; Zafar, A. A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography. *Sol. Energy* **2020**, *208*, 643–651. [[CrossRef](#)]
26. Wang, J.; Gao, D.; Zhu, S.; Wang, S.; Liu, H. Fault diagnosis method of photovoltaic array based on support vector machine. *Energy Sources Part A Recover. Util. Environ. Eff.* **2019**, 1–16. [[CrossRef](#)]
27. Eskandari, A.; Milimonfared, J.; Aghaei, M.; Reinders, A.H. Autonomous monitoring of Line-to-Line faults in photovoltaic systems by feature selection and parameter optimization of support vector machine using genetic algorithms. *Appl. Sci.* **2020**, *10*, 5527. [[CrossRef](#)]
28. Eskandari, A.; Milimonfared, J.; Aghaei, M. Optimization of SVM classifier using Grid Search Method for Line-Line Fault Detection of Photovoltaic Systems. In Proceedings of the 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), Calgary, AB, Canada, 15 June–21 August 2020; pp. 1134–1137.
29. Lazzaretti, A.E.; Costa, C.H.d.; Rodrigues, M.P.; Yamada, G.D.; Lexinoski, G.; Moritz, G.L.; Oroski, E.; Goes, R.E.d.; Linhares, R.R.; Stadzisz, P.C.; et al. A monitoring system for online fault detection and classification in photovoltaic plants. *Sensors* **2020**, *20*, 4688. [[CrossRef](#)]
30. Hussain, M.; Dhimish, M.; Titarenko, S.; Mather, P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renew. Energy* **2020**, *155*, 1272–1292. [[CrossRef](#)]
31. Basnet, B.; Chun, H.; Bang, J. An intelligent fault detection model for fault detection in photovoltaic systems. *J. Sens.* **2020**, *2020*, 6960328. [[CrossRef](#)]
32. Wang, X.X.; Dong, L.; Liu, S.Y.; Hao, Y.; Wang, B. A Fault Classification Method of Photovoltaic Array Based on Probabilistic Neural Network. In Proceedings of the 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3–5 June 2019; pp. 5260–5265.
33. Zhu, H.; Ahmed, S.A.Z.; Alfakih, M.A.; Abdelbaky, M.A.; Sayed, A.R.; Saif, M.A.A. Photovoltaic failure diagnosis using sequential probabilistic neural network model. *IEEE Access* **2020**, *8*, 220507–220522. [[CrossRef](#)]
34. Zhu, H.; Lu, L.; Yao, J.; Dai, S.; Hu, Y. Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. *Sol. Energy* **2018**, *176*, 395–405. [[CrossRef](#)]

35. Sun, E. Small-scale image recognition based on Cascaded Convolutional Neural Network. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; Volume 5, pp. 2737–2741.
36. Moon, W.K.; Lee, Y.W.; Ke, H.H.; Lee, S.H.; Huang, C.S.; Chang, R.F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [[CrossRef](#)]
37. Gao, Z.; Edirisinghe, E.; Chesnokov, S. Image super-resolution using CNN optimized by self-feature loss. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 2816–2820.
38. Ghorbani, R.; Ghousi, R. Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE Access* **2020**, *8*, 67899–67911. [[CrossRef](#)]